FHWA Workshop over the Web for Travel Model Development
Session 4 Homework
Estimation of a Logit Model

Total Time Required: 60-75 Minutes

In this exercise, we will look at the development of a logit mode choice model. The emphasis will be on tracing the "evolution" of a logit model rather than on actually estimating one. This is because logit estimation requires additional software. We will also interpret and analyze the final mode choice model.

We would like to thank Professors Chandra Bhat (UT Austin) and Frank Koppelman[1] (Northwestern University) for providing the models we used in this homework. They estimated these models using survey data from the San Francisco Bay Area.

Please download the MS Excel file "*Homework 4.xls*". The worksheet **Mode Distribution** lists the available transportation modes and also provides details on the market share of each mode for work trips. Go through this table and answer the following questions:

Part 1

1. How do the market shares of modes compare with those in your area?

2. Comment on the average travel times and costs for each mode.

3. What should the auto costs comprise? Remember, we are talking about out-of-pocket costs for each trip that is made using an automobile.

We will now focus on the models themselves. The worksheet **Model Estimation Results** shows results from four versions of the multinomial logit mode choice model. You will note that the models become progressively more elaborate. Each version is commonly called a "specification". This is because model estimation is an iterative process. The modeler needs to "specify" what variables she thinks do a good job of explaining mode choice.

The columns highlighted in yellow are the actual model parameters. The column named "t-stat" helps one judge the statistical significance of these parameters. A t-stat that has an

---

[1] Koppelman, F.S., and C.R. Bhat, "A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models," prepared for the U.S. Department of Transportation, June 2006. http://www.ce.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf

absolute value of more than 1.65 generally indicates statistical significance of the estimated coefficient of the corresponding variable, at the 90% level. At the bottom of each column, are "goodness-of-fit" measures. These measures allow the modeler to compare the relative performance of one model over the other.

Go through the list of variables in each specification and note how the model is progressing from the most simple to the final specification, and answer the following questions:

Part 2

1. What is the dependent variable in this (or any) mode choice model?

2. How do you interpret Model 1? How about Model 2?

3. Model 3 builds on Model 2 by adding an income variable and introducing travel time and cost. Focus on the following questions:

   a. Comment on the signs of the travel time and cost variables.

   b. Calculate the implied value of time (VOT) for this model. (Use in-vehicle time for this calculation.)

   c. Is the income variable significant for the auto (Drive alone and shared ride) modes? How do you interpret the significance (or lack thereof)?

   d. Comment on the negative signs of the coefficients of the income variable for the transit, bike, and walk modes. What are these coefficients negative relative to?

4. Model 4 builds on Model 3 by separating out motorized and non-motorized travel time variables. It also uses separate coefficients for in-vehicle time and out-of-vehicle time components. Focus on the following questions:

   a. Notice that the Out-of-Vehicle travel time is not introduced directly. Instead, it is normalized by the trip distance. Why do you think this might be a good idea?

   b. Notice that the travel cost is not introduced directly. Instead, it is normalized by income. Why do you think this might be a good idea?

   c. Interpret the coefficients of the autos per worker, CBD Dummy, and Employment density. Note the relative magnitudes and signs relative to the coefficients for other modes.

   d. Compute the value of motorized in-vehicle time. Be careful with the cost coefficient here. It is normalized by income!

5. Look at the log-likelihood value at convergence (Row 50), and the rho-squared values (Rows 51 and 52). What do you observe? What does this tell about the relative performance of the models?